Title
Visualizing the Boni dialects with Historical Glottometry

Journal
Journal of Historical Linguistics, 9(1)

ISSN
2210-2116

Author
Elias, Alexander

Publication Date
2019-07-02

Peer reviewed

Visualizing the Boni dialects with Historical Glottometry

Alexander Elias
Leiden University

This paper deals with the historical relations between dialects of Boni, a Cushitic language of Kenya and Somalia. Boni forms the subject of Volume 10 of the *Language and Dialect Atlas of Kenya* (Heine & Möhlig 1982). Heine presents evidence for three subgroups within Boni, as well as several areas of convergence between dialects belonging to different proposed subgroups. In reviewing his evidence, I find that two of the three splits are not supported by the data, and therefore his conclusions on convergence must also be reinterpreted. Given the presence of numerous intersecting isoglosses, the tree diagram is an inappropriate model for describing the relations between Boni dialects, and I turn to Historical Glottometry (Kalyan & François 2018) to provide a visualization of the data.

Keywords: Boni dialectology, Historical Glottometry, wave model

1. Introduction

1.1 The Boni dialect group

Boni is a dialectally diverse Cushitic (Afroasiatic) language spoken mainly in Lamu District of northern Kenya, with some speakers of the northernmost dialect Kilii over the border in southern Somalia. The 2009 Kenya census counted Boni 7,600 speakers (Kenya Census 2009; reported in Simons & Fennig 2017).

The anthropological information cited below is taken from Stiles (1988; passim). Boni speakers are traditionally hunter-gatherers inhabiting lowland forest areas slightly inland from the coast. The potentially derogatory name "Boni" is proposed to derive from a southern Somali word *bon*, referring to a lower caste of hunter-gatherers in an unequal economic relationship with pastoralist groups. This type of social structure is attested elsewhere in the area (the Midgaan of northern Somalia and the Ribe along the middle Juba River), but the Boni are the only ones with their own language. The Boni of Kenya refer to themselves as "Aweer", but the Aweer autonym excludes the northernmost Kilii. Thus, following the authors of the *Language and Dialect Atlas of Kenya* (*LDAK*; Heine & Möhlig 1982), I will use "Boni" to refer to the group as a whole, and "Aweer" to refer to the subset of dialects which excludes the northernmost Kilii dialect.

Boni is most closely related to Somali and Rendille, together forming the Sam subgroup of Eastern Cushitic. Heine places the pastoralist Proto-Sam around 0-300 AD, and proposes that they migrated south-west to the coast from a homeland in inland northern Kenya. They spread along the coast, and the group that became the Boni gave up animal

husbandry as an adaptation to the forest environment they found there (Heine 1978; passim). In more recent times, the Boni transitioned from a hunter-gatherer lifestyle to become settled agriculturalists (Heine & Möhlig 1982; p12).

The most divergent dialect of Boni is Kilii, which Heine states is not "immediately intelligible" with the others (Heine & Möhlig 1982; p12). Therefore, it may be justifiable to refer to Kilii as a separate language, but in this paper I will refer to it as a dialect of Boni. The remaining dialects, on the other hand, are not strongly differentiated and remain mutually intelligible. Stiles (1988; p43) dates the split between Kilii and the others to the 16th century, writing the following on the arrival of the Oromo at that time:

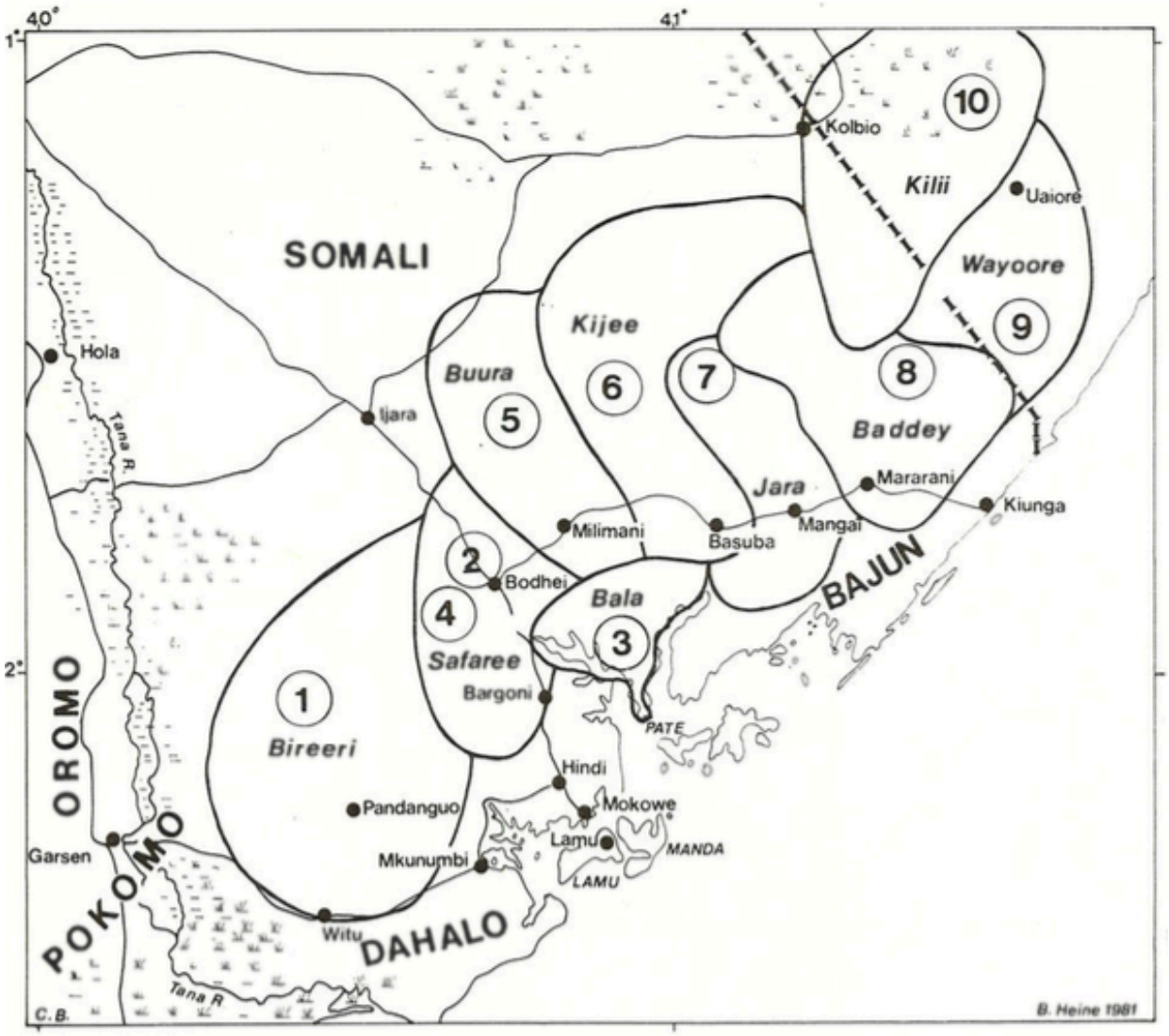
> The arrival of the Oromo changed the situation. Whatever organization that previously existed was broken down and people began to migrate south and to the off-shore islands ... The Oromo must have been in the area as my informants claim that at this time the Boni would set upon Bajunis and sell them and their property to the Oromo. Before this time the Boni say that the Oromo were not in Shungwaya. The Boni who remained with the Bajunis when they fled to the islands are today known as the Wakatwa; the Boni who stayed in the southern Somalia forests are called Kilii, and the Boni who later associated with the Somalis are called Garra ... The dialect differentiation between Kilii and the other Boni dialects which Heine (1982) notes probably occurred at this time, i.e. in the 16th century.

Thus, the time depth of differentiation of the Boni dialects is not very great, and this is reflected in their many similarities. Importantly, Stiles (1988) proposes a concrete historical event, the invasion of the Oromo, which led to a physical split in the Boni community, with the Kilii dialects remaining in southern Somalia and the remaining dialects migrating south to their present locations in northern coastal Kenya. The migrating group ("Aweer" in this paper) remained relatively cohesive as they moved south to their current area in Lamu District and entered into relations with the Oromo people there. The Kijee are said to be made up of Oromo people who joined the Boni, and the word *Kijee* means "Oromo" in Boni (Stiles 1988; p43-44).

1.2 Data source

The data in this paper is drawn entirely from Volume 10 of the *LDAK*, which deals specifically with the Boni dialects. Prior to this publication, it was known that Boni was dialectally diverse but work had focused only on the Bireeri dialect (Heine 1977, Sasse 1979). This volume of *LDAK* presents the results of fieldwork carried out by Heine on the dialects Jara and Kilii, with comparative data from the other dialects. The first section is a compact overview of the phonology and morphology of Jara and Kilii, and the second section deals with relations between dialects and reconstruction of Proto-Boni. The following map (Fig. 1) shows the layout of the traditional Boni territorial divisions along the coast of Kenya and Somalia. Data from 6 of these 10 divisions is presented in the *LDAK* - Bireeri (1), Safaree (4), Kijee (6), Jara (7), Baddey (8) and Kilii (10):

Figure 1: Map of traditional Boni territorial divisions (reproduced from Heine & Möhlig 1982; p14). The dotted line in the top right corner is the border between Kenya and Somalia



Heine & Möhlig (1982; p17-18) list 13 people as their main informants. I list them below with their initials, sex, age, place of birth/residence and the name of their native dialect, when the information is included. No further details are provided on how the data was collected.

Table 1: List of informants in Heine & Möhlig (1982; p17-18)

Initials	Sex	Age	Residence	Dialect
AD	male	25	Basuba	Bala
KB	male	32	Mangai	Kijee
BB	male	35	Badada	Kilii
FH	female	18	Mangai	not listed
AM	female	19	Mangai	not listed
JC	male	18	Mangai	not listed

OJ	male	16	Mangai	Jara
OM	male	18	Mangai	Kijee
MB	male	11	Bargoni	Safaree
SA	male	13	Bargoni	Bala
HB	male	32	Bargoni	Safaree
AK	male	33	Bargoni	Buura
GW	male	35	Bargoni	Kijee

The Boni dialects are typologically almost identical to each other. The basic word order is SOV and modifiers usually follow their head. They have large consonant inventories, which include ejective voiceless and implosive voiced stops (but the latter are not present in all dialects). All Boni dialects show vowel harmony between a set of tense and lax vowels, and they have a two-tone system where each syllable takes either a high or low tone. They show complex interactions between gender and number (termed "gender polarity", following Meinhof 1912) often seen in Cushitic languages.

Boni speakers are a minority in both Kenya and Somalia, and as a result they are often competent in a local lingua franca. A majority of Boni speakers of both genders are competent in Swahili (Bantu), the lingua franca of East Africa. This has introduced loanwords from Bantu into all Boni dialects to varying degrees. There are also two Cushitic languages commonly spoken by Boni speakers: Oromo and Somali. Most male Boni speakers are competent in one of these two languages, and the influence of Oromo is clearest in the peripheral dialects (Bireeri and Kilii). Although English is the national language of Kenya, less than 3% of Boni men were competent in English at the time of writing of the *LDAK* (Heine & Möhlig 1982; p12).

1.3 Theoretical background

Not all language groups are equally well represented by a tree diagram. The tree diagram is most appropriate when the language group has undergone a series of splits, whereby one group of speakers loses contact with the rest, and both groups undergo a period of exclusively shared innovation in isolation from each other. These splits correspond to nodes on a tree diagram and are understood to reflect real events of societal splitting in the ancestral speech community. In this idealized scenario, isoglosses will never intersect, and distinguishing between vertical and horizontal transmission is usually not problematic. However, this pattern of strictly non-intersecting innovations will only hold in the ideal case. In cases where there were no abrupt splits in the speech community, there can indeed be intersecting isoglosses which complicate construction of a tree diagram. Furthermore, since nodes on a tree diagram correspond to actual splitting events, the tree diagram has nothing to represent if the community did not undergo any splits. Overreliance on the tree model leads to the proposal of unrealistic splits, and often requires the linguist to sweep certain problematic data under the rug in order to arrive at a strictly nested set of isoglosses.

The problematic reasoning which the tree model can lead to is found in Heine & Möhlig (1982). For instance, Heine proposes the subgroup "Central Aweer" consisting of Safaree, Jara and Kijee based on the fact that the Proto-Boni distal demonstrative /*-*óó*h*e*/ changes to /*-*óó*h*o*/ in those three dialects (Heine & Möhlig 1982; p81). Heine then notes

that a number of innovations group the dialects of Bireeri, Safaree and Kijee together, to the exclusion of Jara. These innovations include the replacement of the Proto-Boni negative perfect prefix /*mə-/ with /húú-/ (< /*hááb-/) and the change of /*b/ > /w/ word-internally, at least in some words (Heine & Möhlig 1982; p85). I have dubbed this the "Southern Convergence Area" in this paper, whereas Heine simply referred to the "North-Central isogloss". Heine is agnostic as to whether this set of innovations reflects a genetic or areal effect, saying its "historical significance remains to be investigated" (Heine & Möhlig 1982; p85). However, it is implicitly assumed by the tree model that any non-genetic group must arise by later convergence, hence the choice of name. The evidence provided in favor of Central Aweer (one irregular sound change in a demonstrative) is in fact weaker than that adduced for the Southern Convergence Area (one morphological innovation and one semi-regular sound change), but in the process of constructing a nested set of innovations, Heine was forced to discard the evidence in favor of the Southern Convergence Area and treat it as an unexplained phenomenon.

This paradox arises mainly from the impossibility of distinguishing vertical and horizontal transmission in such a closely-knit group of dialects. The tree diagram fails to accurately capture the linguistic history of groups which differentiate while maintaining contact and mutually intelligibility with closely related speech varieties. Such a situation will lead to the formation of a dialect chain at first, and if the process of differentiation continues, a linkage (as defined in Ross 1988). In a dialect chain or linkage, intersecting isoglosses are an empirical fact which simply cannot be accounted for under the tree model. The wave model provides an alternative which can account for these intersecting isoglosses and provide a principled way to avoid Heine's paradox in the first place: in the wave model, linguistic change is viewed as a set of ripples spreading in a pond (an image put forth by the first proponent of the wave model, Schmidt 1872). A change begins at an innovatory center and spreads to neighboring speech varieties. The change can continue spreading as long as there is contact and some level of intelligibility or at least bilingualism between neighboring populations. Each change can spread over a different portion of the whole speech area, resulting in the pattern of intersecting isoglosses commonly observed in dialect chains. This is a much more natural and realistic account of differentiation in cases where there is no reason to suspect that any major splitting events have separated speakers from one another. In addition, the wave model has the advantage that tree-like language groups are simply a special case in which isoglosses happen not to intersect.

The reason that the tree model remains so dominant over the wave model is that the tree model provides an intuitive visual representation in the form of a tree diagram. The wave model has seen numerous proposals for visualizations, including trees modified with double lines (Ross 1988), isogloss maps (Anttila 1989), and NeighborNets (Bryant et al. 2005). These suffer from various drawbacks of either informativity or readability. One recent attempt to improve on these methods is the glottometric diagram in the framework of Historical Glottometry (HG). This is an ongoing project led by Alexandre François and Siva Kalyan, laid out in a series of recent publications (François 2014, François 2017, Kalyan & François 2018). I have chosen to recast the data contained in the *LDAK* in the framework of HG, and to present the findings in the form of a glottometric diagram. Boni makes a good case study for developing methods of visualizing non-tree-like language groups because the conditions for the overlapping spreading of innovations are present: mutual intelligibility and continued contact.

2. Review of Heine's Evidence

In this section I will present the reconstructed phonology of proto-Boni and the evidence presented by Heine in support of the various groupings within Boni that he identifies. He first presents the divergent changes that he takes as evidence for subgroups within Boni, followed by convergent changes which intersect the subgroup boundaries. Based on the evidence provided, I conclude that only one of the three proposed subgroups is strongly supported by linguistic (and historical) evidence. I assign numbers to each of the 25 innovations he identifies, and this body of evidence will then form the input for the creation of a glottometric diagram.

2.1 Proto-Boni Phonology

According to Heine, proto-Boni phonology is unproblematic to reconstruct because the modern dialects agree quite closely with each other (Heine & Möhlig 1982; p71). He reconstructs the following inventories of consonant and vowel phonemes:

Table 2: Reconstructed Proto-Boni consonant inventory (Heine & Möhlig 1982; p71)

	Labial	**Dental/Alveolar**	**(Alveo)palatal**	**Velar**	**Uvular**	**Glottal**
Stop	*p *b	*t *d	*c *ɟ	*k *g	*q	*ʔ
Ejective		*tʼ	*cʼ	*kʼ		
Implosive		*ɗ	*ʄ	*ɠ		
Fricative	*f	*s	*ʃ			*h
Trill		*r *ɽ				
Lateral		*l				
Nasal	*m	*n	*ɲ	*ŋ		
Glide	*w		*j			

Table 3: Reconstructed Proto-Boni vowel inventory (Heine & Möhlig 1982; p71)

	Front		**Central**		**Back**	
	Lax	**Tense**	**Lax**	**Tense**	**Lax**	**Tense**
Close	*i	**i*			*u	**u*
Mid	*e	**e*	*ə	**ə*	*o	**o*
Open			*a	**a*		

Each vowel comes in a lax-tense pair, and there is a system of vowel harmony where lax vowels can assimilate to tense ones across morpheme boundaries. Tense vowels are spelled in italics in this paper (/*i e a o u*/), following the usage of the *LDAK*. Phonetically, the tense vowels are close to the cardinal vowels while the lax vowels are more centralized. Vowels can appear long or short, and the long vowels are indicated by doubled spelling. It is also possible to reconstruct a two-tone system where each syllable can take a high or low tone.

In this paper, the high tone is indicated by an acute accent (/V́/) while the low tone is left unmarked.

2.2 Proposed splits

2.2.1 *Aweer*

The first proposed split is between Kilii and Aweer (consisting of the other 5 dialects: Bireeri, Safaree, Jara, Kijee and Baddey) (Heine & Möhlig 1982; p78). This evidence is strong enough to justify Aweer as a valid subgroup, created by a genuine split between Kilii and Proto-Aweer. The linguistic evidence of a split is corroborated by historical evidence of an Oromo invasion in the 16th century (see section 1.1; and Stiles 1988) which caused the Proto-Aweer speakers to flee south, while the Kilii speakers stayed in place. The following 9 innovations support this split. After each innovation, the page number where it is introduced is listed, as well as the page numbers of each example in the wordlist at the end:

Innovation #1: Proto-Boni *nt > Proto-Aweer *t, Kilii /n/

Table 4: Illustration of innovation #1 (Heine & Möhlig 1982; p78, 135)

English	Proto-Boni	Proto-Aweer	Kilii
'to vomit'	*məntah-	*mətah-	manah-

Innovation #2: Proto-Boni *dt > Proto-Aweer *dd, Kilii /tt/

Table 5: Illustration of innovation # 2 (Heine & Möhlig 1982; p78, 106, 115)

English	Proto-Boni	Proto-Aweer	Kilii
'fingernail, claw'	*idtíʔi	*iddíʔi	ittíi
'leg'	*ídtə	*íddə	íttə

Innovation #3: Proto-Boni *m > Proto-Aweer *w, *j between unlike vowels, deleted between like vowels

Table 6: Illustration of innovation #3 (Heine & Möhlig 1982; p79, 102, 99, 90)

English	Proto-Boni	Proto-Aweer	Kilii
'to destroy, spoil'	*humeej-	*huwej-	humeej-
'to come (impf.)'	*-imaad-	*-ijaad-	-imaad-
'to be angry'	*umuʃooɗ-	*uuʃooɗ-	umusooʔ-

Innovation #4: Proto-Boni *m > Proto-Aweer *ŋ morpheme-finally

Table 7: Illustration of innovation #4 (Heine & Möhlig 1982; p79, 104, 107)

English	Proto-Boni	Proto-Aweer	Kilii
'to eat'	*-aham-	*-ahaŋ-	-aham-
'food made of flour'	*məsím-ə	*məsíŋə	məsímə

Innovation #5: Proto-Boni *t’ > Proto-Aweer *d, Kilii /t/

Table 8: Illustration of innovation #5 (Heine & Möhlig 1982; p79, 108, 116)

English	Proto-Boni	Proto-Aweer	Kilii
'forest'	*t'úur	*dúur	túur
'maggot'	*t'iʔírə	*diʔírə	tihírə

Innovation #6: Proto-Boni *kʃ > Proto-Aweer *ʃ, Kilii /ks/
Table 9: Illustration of innovation #6 (Heine & Möhlig 1982; p79, 134)

English	Proto-Boni	Proto-Aweer	Kilii
'topi'	*h*a*wk*ʃéé*	*h*a*w*ʃée*	h*o*ks*é*

Innovation #7: Proto-Boni *mb > Proto-Aweer *b
Table 10: Illustration of innovation #7 (Heine & Möhlig 1982; p79, 117)

English	Proto-Boni	Proto-Aweer	Kilii
'marabou stork'	*b*áá*mb*o*	*b*áá*b*o*	*b*áá*mb*o*

Innovation #8: Low vowels assimilate to a following non-low vowel in some words
Table 11: Illustration of innovation #8 (Heine & Möhlig 1982; p79, 90, 91, 107)

English	Proto-Boni	Proto-Aweer	Kilii
'ant species'	*ʃarííra	*ʃirííra	ʃarííra
'lower arm'	*t*a*g*ó*g	*t*ó*g*ó*g	t*a*g*ó*g
'flower, blossom'	*barúúrə	*burúúrə	barúúrə

Innovation #9: Vowels are rounded before *b or *w (in some words)
Table 12: Illustration of innovation #9 (Heine & Möhlig 1982; p79, 97, 99)

English	Proto-Boni	Proto-Aweer	Kilii
'to catch, seize'	*qabəɗ-	*ɠobəɗ-	k'abaʔ-
'civet cat'	*símáad	*suwáad	sim*áa*d

2.2.2 *South Aweer*

The second proposed split is between Baddey and South Aweer (consisting of Bireeri, Safaree, Jara and Kijee) (Heine & Möhlig 1982; p79). The evidence for this as a genuine split is weak, and therefore the existence of a South Aweer subgroup is not strongly supported. Only one piece of unproblematic evidence is adduced:

Innovation #10: Proto-Boni *w is deleted between two /a/'s in Proto-South Aweer
Table 13: Illustration of innovation #10 (Heine & Möhlig 1982; p81, 90)

English	Proto-Boni	Proto-South Aweer	Baddey
'harmless edible animal'	*buʃáwa	*buʃáa	buʃáwa

Two more pieces of evidence are adduced by Heine in support of the subgroup South Aweer, but one of them is in fact a shared retention (Proto-Boni *ɟ > /j/ in Baddey, but is retained in

South Aweer: Heine & Möhlig; p81). Furthermore, the regular shift of PB *ɠ > *ʔ in South Aweer which he cites as evidence (Heine & Möhlig 1982; p79) is questionable because the same change also affects Kilii. Kilii may well have innovated this independently, since there is a parallel change of PB *ɗ > /ʔ/ in Kilii as well (Heine & Möhlig 1982; p78). If it is the case that these two innovations are the same by coincidence, then this is a second shared innovation in support of South Aweer. However, I have left it out of the final analysis, given the uncertainty surrounding it.

2.2.3 *Central Aweer*

The third proposed split is between Bireeri and Central Aweer (Safaree, Jara and Kijee) (Heine & Möhlig 1982; p81). The evidence for this as a genuine split is also weak, and therefore the existence of a Central Aweer subgroup is not strongly supported. It consists of a single piece of unproblematic evidence:

Innovation #11: Vowel assimilation in distal marker

Table 14: Illustration of innovation #11 (Heine & Möhlig 1982; p81, 75)

English	Proto-Boni	Proto-Central Aweer	Bireeri
'that, distal'	*-*óó*h*e*	*-*óó*h*o*	*missing*

The other evidence offered is a set of consonant cluster simplifications which occur only in Bireeri and thus in fact constitute a shared retention in Proto-Central Aweer.

2.3 Proposed convergence areas

In the next section, Heine presents evidence for changes which cut across the boundaries of the genetic groups previously established. The implicit hypothesis of the tree model is that these must have arisen by areal convergence, and this is indeed how Heine presents the peripheral convergence with Oromo (section 2.3.4) and the Central Convergence Area (2.3.3). As for Southern Convergence Area (2.3.1) and South-Central Convergence Area (2.3.2), these are newly coined names; Heine refers to them as the "South-Central isogloss" and "North-Central isogloss" respectively (Heine & Möhlig 1982; p85). He is agnostic as to whether they are the result of genetic splitting or areal convergence, but I have listed them here as convergence areas because that is the implicit position taken, assuming they are non-genetic groups.

2.3.1 *Southern Convergence Area*

The neighboring pair of southern dialects Bireeri and Safaree form the Southern Convergence Area (bounded by the "South-Central isogloss" in Heine & Möhlig 1982; p85). They share a common morphological innovation, whereby the negative imperative plural suffix is replaced by the affirmative imperative suffix. They also have 4 lexical innovations in common:

Innovation #12: PB *-náha 'neg. imperative pl.' is lost in favor of affirmative *-níja

(Heine & Möhlig 1982; p85, 75-76)

4 Lexical Innovations:

Table 15: Lexical innovations #13-16 affecting the Southern Convergence Area (Heine & Möhlig 1982; p82, 132, 126, 134, 137)

English	Proto-Boni	Jara, Baddey	Kilii	Kijee	Bireeri	Safaree
#13 'tamarind'	*múkaj	múkaj	múkaj	*missing*	**róɠa**	**róka**
#14 'sheep'	*aléeŋ	aléeŋ	aléeŋ	*missing*	**hoolá**	**hoola**
#15 'to touch'	*taar-	taar-	taab-aʔ-	*missing*	**k'ol-, taar-**	**k'ol-, ɠol-**
#16 '2PL'	*ado	ado	ado	ado	**isanə**	**is(i)no**

2.3.2 *South-Central Convergence Area*

These dialects are roughly those which lie south of the Dodori river: Bireeri, Safaree and Kijee. They are bounded by Heine's "North-Central isogloss" (Heine & Möhlig 1982; p85). There are at least 2 innovations which they share:

Innovation #17: Proto-Boni *mə- 'neg. perfect' > SCCA /húú-/
(Heine & Möhlig 1982; p85, 76)

Innovation #18: Proto-Boni *b > SCCA /w/ word internally

Table 16: Illustration of innovation #18 (Heine & Möhlig 1982; p85, 97, 100)

English	Proto-Boni	SCCA	Jara	Baddey	Kilii
'to catch, seize'	*qabəɗ-	owɗ-	obəɗ-	ɠobɗ-	k'abaʔ-
'to curse'	*habaaɽ-	hawaar-	habaar-	*missing*	habaaɽ-

2.3.3 *Central Convergence Area*

This area covers the dialects Safaree, Kijee, Jara and Baddey (all except the peripheral dialects Bireeri and Kilii) (Heine & Möhlig 1982; p82). There are 5 shared innovations cited for this group of dialects, including two instances of shared loans from Bantu:

Innovation #19: Proto-Boni *b > CCA /f/ word-finally (in some words)

Table 17: Illustration of innovation #19 (Heine & Möhlig 1982; p84, 131)

English	Proto-Boni	CCA	Bireeri	Kilii
'small sword'	*gólhob	gólof	gólob	gólhob

Innovation #20: PB *lk > CCA /rk/

Table 18: Illustration of innovation #20 (Heine & Möhlig 1982; p84, 134)

English	Proto-Boni	CCA	Bireeri	Kilii
'teeth'	*ilkə	irki, irkə	ilkə	ekə́

Innovation #21: PB *ɠt > CCA /gt/, Bireeri, Kilii /tt/

Table 19: Illustration of innovation #21 (Heine & Möhlig 1982; p84, 101, 128)

English	Proto-Boni	CCA	Bireeri	Kilii
'darkness'	*múɠtə	múgti, múgtə	múttə	múʔuttə
'sisal-like plant'	*ɗeɠté	ɗégtə	*(hargéi)*	etté

2 Lexical innovations (shared loans from Bantu):

Table 20: Lexical innovations #22-23 affecting the Central Convergence Area (Heine & Möhlig 1982; p84, 95, 111)

English	Proto-Boni	CCA	Bireeri	Kilii
#22 'blacksmith'	*tumal	fuundə, fúúndi	túma	tumáal
#23 'hare'	*hileesə	kituŋgúe	híleesə	hiléésə

2.3.4 *Peripheral convergence with Oromo*

This is a fundamentally different type of convergence which affects the peripheral dialects Bireeri and Kilii most. They are both under particularly strong influence from Oromo, according to Heine & Möhlig (1982; p84-85). They are not directly converging with each other, but rather converging on the same target, leading to shared innovations between them. Heine lists two Oromo loanwords shared exclusively by the peripheral dialects as evidence for this process:

2 Lexical innovations (shared loans from Oromo):

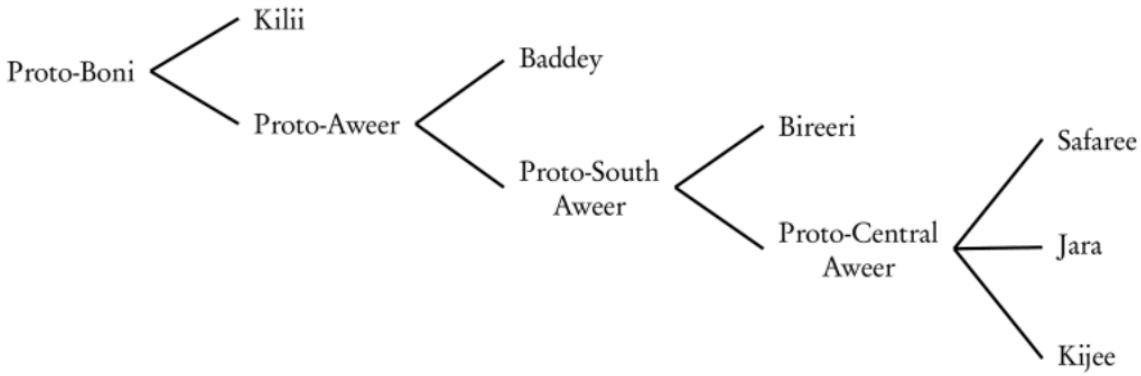
Table 21: Lexical innovations #24-25 affecting the peripheral dialects Bireeri and Kilii (Heine & Möhlig 1982; p84, 99, 118)

English	Oromo	Bireeri	Kilii	Other
#24 'cloud'	duumans, dúmes	duumánsə	duuməssə	*missing*
#25 'vervet monkey'	k'amaleʔ	k'amála	k*a*m*á*l*e*	tabáábu, *missing*

2.4 Assessment of evidence

Heine proposes three nested subgroups within Boni: Aweer, South Aweer and Central Aweer. He then proposes the existence of convergence zones to explain the fact that many isoglosses cut across the boundaries of the proposed subgroups. Heine's view of the splits within Boni is summed up by the tree below. Heine does not summarize his proposal on the genetic splits in a tree diagram, so I have created the following tree which represents his subgrouping proposal:

Figure 2: Heine's view of the splits within Boni (my interpretation)



However, given the strength of the evidence presented, only the existence of the first subgroup (Aweer) is well-supported. This includes both linguistic and historical evidence of a split in the Proto-Boni community due to the invasion of the Oromo, dated to approximately the 16th century (Stiles 1988). The evidence for the proposed South Aweer and Central Aweer groups consists of a single innovation in both cases (possibly two in the case of South Aweer), and neither is strong enough to confidently identify a genuine splitting event in the history of Boni. Furthermore, no major historical event like the arrival of the Oromo can be identified which could potentially have caused such splits. Therefore, Heine's analysis that Proto-Boni split into Kilii and Proto-Aweer is well-supported, but there is not strong enough evidence to posit tree-like subgroups below the level of Proto-Aweer.

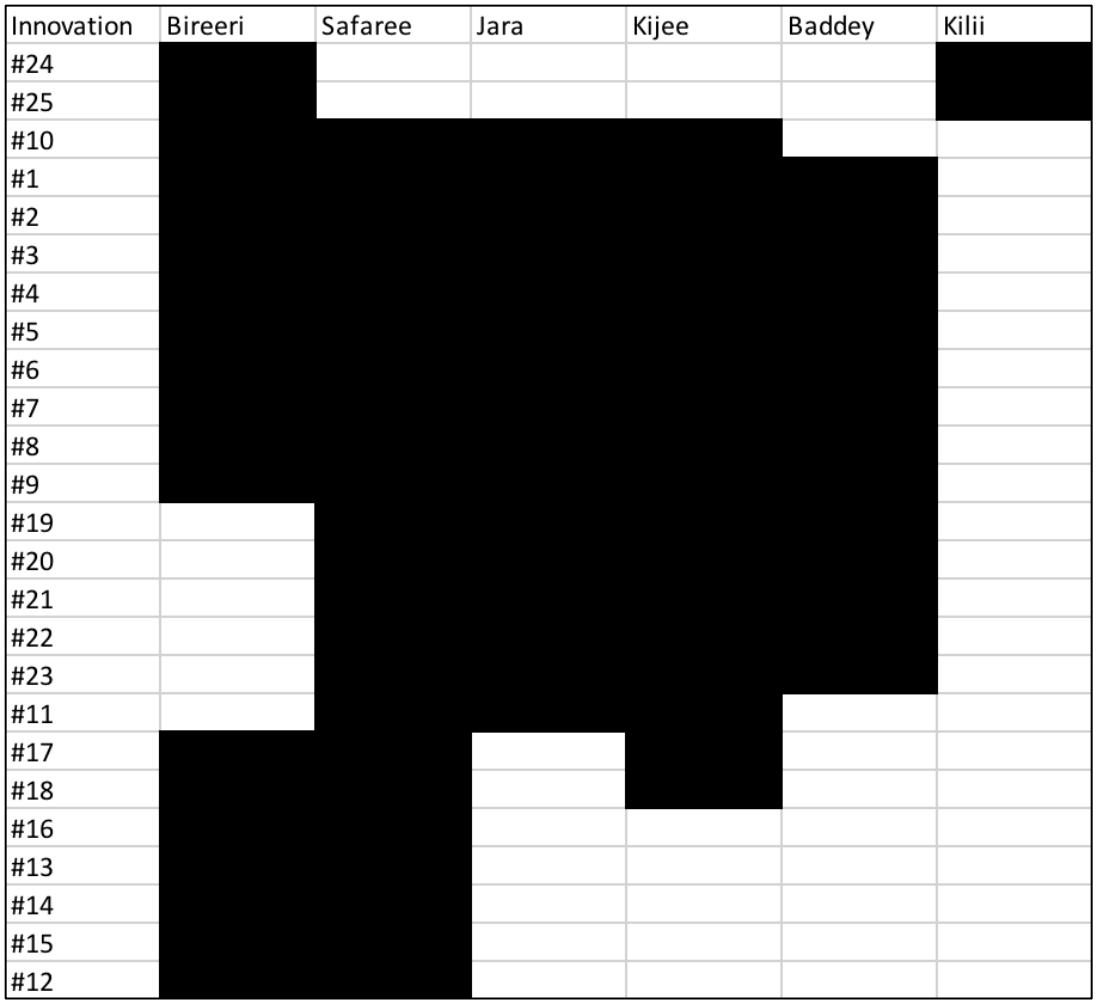
3. **Glottometric Analysis**

I conclude two things from reviewing Heine's evidence. Firstly, I conclude that there was indeed a genuine split between speakers of Kilii and Proto-Aweer. Secondly, I conclude that below the level of Proto-Aweer there is no evidence for genuine splits in the Boni-speaking community. Instead, the differences between modern dialects are best understood as the result of wave-like processes of language differentiation, enabled by the continued mutual intelligibility and contact between these dialects. Given that there is only one well-supported split, a tree diagram would contain only a single node at Proto-Aweer and would not be the most informative method of representing the history of the Boni dialects. It is desirable to represent something not only about the splits, but also the overlapping waves of change which have targeted different sets of dialects. I will now explain the methodology behind creating a glottometric diagram using an example calculation.

3.1 Methodology

The database of innovations which formed the input for the diagram in this case looks like this, when cells are shaded in to represent their participation in an innovation:

Figure 3: Database of shared innovations among Boni dialects; shading indicates participation in an innovation



Innovation	Bireeri	Safaree	Jara	Kijee	Baddey	Kilii
#24	■					■
#25	■					
#10	■	■	■	■		
#1	■	■	■	■	■	
#2	■	■	■	■	■	
#3	■	■	■	■	■	
#4	■	■	■	■	■	
#5	■	■	■	■	■	
#6	■	■	■	■	■	
#7	■	■	■	■	■	
#8	■	■	■	■	■	
#9	■	■	■	■	■	
#19		■	■	■	■	
#20		■	■	■	■	
#21		■	■	■	■	
#22		■	■	■	■	
#23		■	■	■	■	
#11		■	■	■		
#17	■	■		■		
#18	■	■		■		
#16	■	■				
#13	■	■				
#14	■	■				
#15	■	■				
#12	■	■				

From this database, it is possible calculate the following values for any set of languages or dialects (as defined in Kalyan & François 2018):

- ε = number of exclusively shared innovations
- p = number of supporting innovations
- q = number of conflicting innovations

The concept of an exclusively shared innovation is already familiar. A supporting innovation is one that affects all the languages or dialects under investigation, but may affect others as well. The important thing is that all the languages or dialects in question innovated together. A conflicting innovation is one where at least one member of the set of languages or dialects participates in innovations with at least one language outside the set. However, the conflicting innovation cannot target all members of the set in question. These conflicting innovations support a different subgrouping and undermine the strength of the evidence for any given subgroup. Taking the set of dialects Bireeri and Safaree as an example (the Southern Convergence Area according to Heine), it is possible to observe that they share 5

exclusive innovations (# 12-16). A further 12 innovations affect Bireeri and Safaree as well as other dialects, for a total of 17 supporting innovations (since exclusively shared innovations count as supporting innovations as well). Finally, Bireeri and Safaree have 8 conflicting innovations (11, 19-25) where one of the pair innovates with some other group of dialects instead. Thus, the values for the Southern Convergence Area are ε = 5, p = 17, q = 8.

Based on these values, the two final quantities which will serve in the creation of the glottometric diagram can be computed. The first is cohesiveness (κ) and it measures the tendency of a set of languages or dialects to innovate together rather than separately, on a scale from 0 to 1. It is equal to the number of supporting innovations over the total number of relevant innovations (that is, both supporting and conflicting ones; see Kalyan & François 2018):

κ = (supporting innovations) / (total relevant innovations)
= p / (p + q)
= 17 / (17 + 8)
= 0.68

Note that under the tree model, all subgroups must receive cohesiveness 1 because conflicting innovations are not recognized as valid. Cohesiveness is typically high in a tree-like subgroup (one which has differentiated through splitting and isolation) and low in a non-tree-like subgroup or dialect chain situation.

The second quantity to be calculated is subgroupiness (*s*). This is simply equal to the number of exclusively shared innovations times the cohesiveness of the set (as defined in Kalyan & François 2018):

ς = (cohesiveness) * (exclusively shared innovations)
= κ * ε
= 0.68 * 5
= 3.4

Subgroupiness is a way of weighting the quantity of the evidence in favor of a subgroup by the quality of that evidence. Exclusively shared innovations are the only acceptable source of evidence in normal linguistic classification, so this method of analysis is orthodox in that regard. However, it adds the recognition that some language groups behave less cohesively than others, and penalizes them as a result. Note that subgroupiness, unlike cohesiveness, has no maximum value, and a larger database with more innovations will lead to larger subgroupiness scores.

When constructing a database of innovations, all types of shared innovations are acceptable. This includes sound changes (both regular and irregular), as well as innovations in morphology, syntax, semantics, and lexicon. Normally, these types of innovations are not taken to be equally indicative of subgrouping. Shared irregularities in morphological paradigms and uncommon sound changes are usually given great weight in language classifications, on the basis that they are less likely to spread through horizontal transmission or be replicated by chance. However, this often turns into a blunt tool for choosing which data is to be used to establish a "genuine subgroup", and which can be safely

discarded. There is no need to discard common sound changes or shared loanwords in the construction of the database, because in the wave model, all innovations spread horizontally. The fact of a common change (like /s/ > /h/) spreading through contact is no less indicative of convergence and shared innovation than a rare sound change. Of course, instances of random parallel innovation, where the same innovation occurs twice by chance, are still unacceptable as subgrouping evidence.

One solution, which would preserve the intuition that rare innovations are more solid indicators of language history than common changes or lexical loans, would be to weight innovations differently according to their type and rarity. However, it would be crucial to avoid the types of intuitive judgments about "common" and "uncommon" changes which we are taking such pains to avoid. In order to do this in an objective way, a large database of innovations from around the world would need to be compiled, but it would be extremely hard to ensure that this database is representative. In view of the difficulties of implementing such a weighting system correctly, it is preferable to simply stick to unweighted values. More information on the possibility of weighting, and reasons for not doing it, can be found in François (2014; p176-177).

3.2 Results

The full table of subgroupiness and cohesiveness values calculated for each group is given below. These values can be calculated from the database of innovations given in Figure 6 above. The sets of dialects are listed in order of subgroupiness, and only groups which received a subgroupiness score greater than 0 are included:

Figure 4: Cohesiveness and subgroupiness scores for all groups which have at least one exclusive innovation

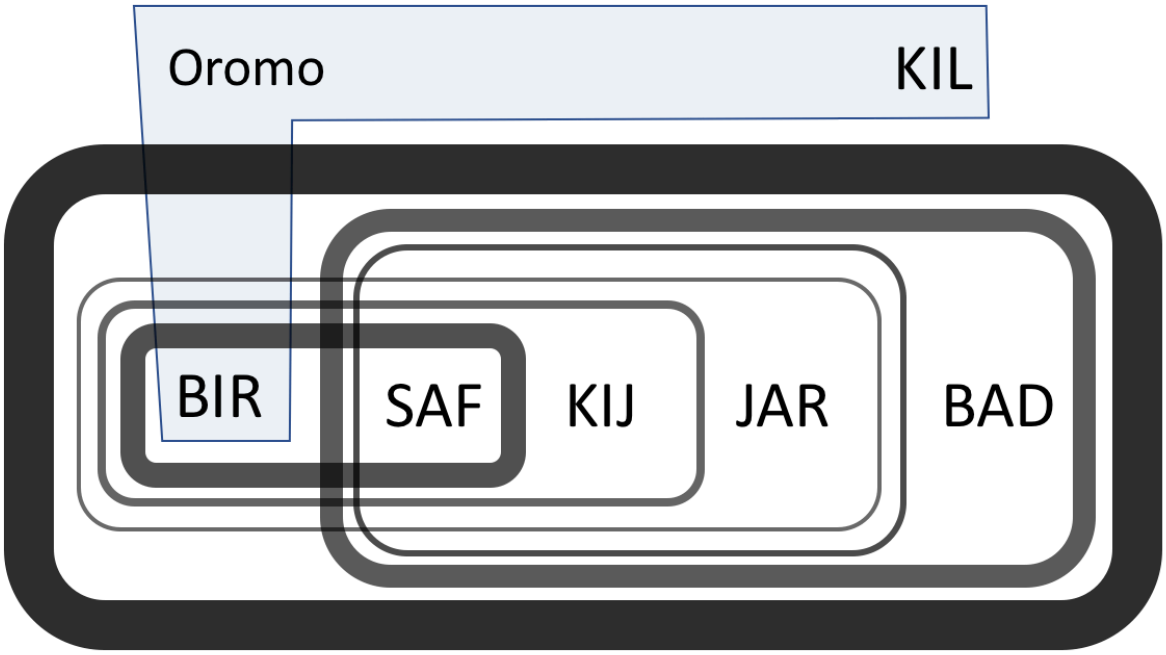
	Exclusive Innovations	**Supporting Innovations**	**Conflicting Innovations**	**Cohesiveness**	**Subgroupiness**
Aweer	9	9	2	0.82	7.38
SCA	5	17	8	0.68	3.4
CCA	5	14	8	0.64	3.2
SCCA	2	12	8	0.6	1.2
CA	1	16	7	0.7	0.7
SA	1	10	7	0.59	0.59
Peripheral	2	2	17	0.11	0.22

Guide to names of dialect groups:

Aweer = Bireeri, Safaree, Jara, Kijee, Baddey
SCA = Bireeri, Safaree (Southern Convergence Area)
CCA = Safaree, Jara, Kijee, Baddey (Central Convergence Area)
SCCA = Bireeri, Safaree, Kijee (South Central Convergence Area)
CA = Safaree, Jara, Kijee (Central Aweer)
SA = Bireeri, Safaree, Jara, Kijee, Baddey (South Aweer)
Peripheral = Bireeri, Kilii

In order to go from these raw values to a glottometric diagram, it is necessary to arrange the dialects in a convenient layout, then surround the various subgroups with lines representing bundles of isoglosses. The thickness of the encircling line is proportional to subgroupiness, while its darkness is proportional to cohesiveness. It can be very tricky to figure out the best configuration, since there is currently no automated way of generating these diagrams. The result is shown in the glottometric diagram below:

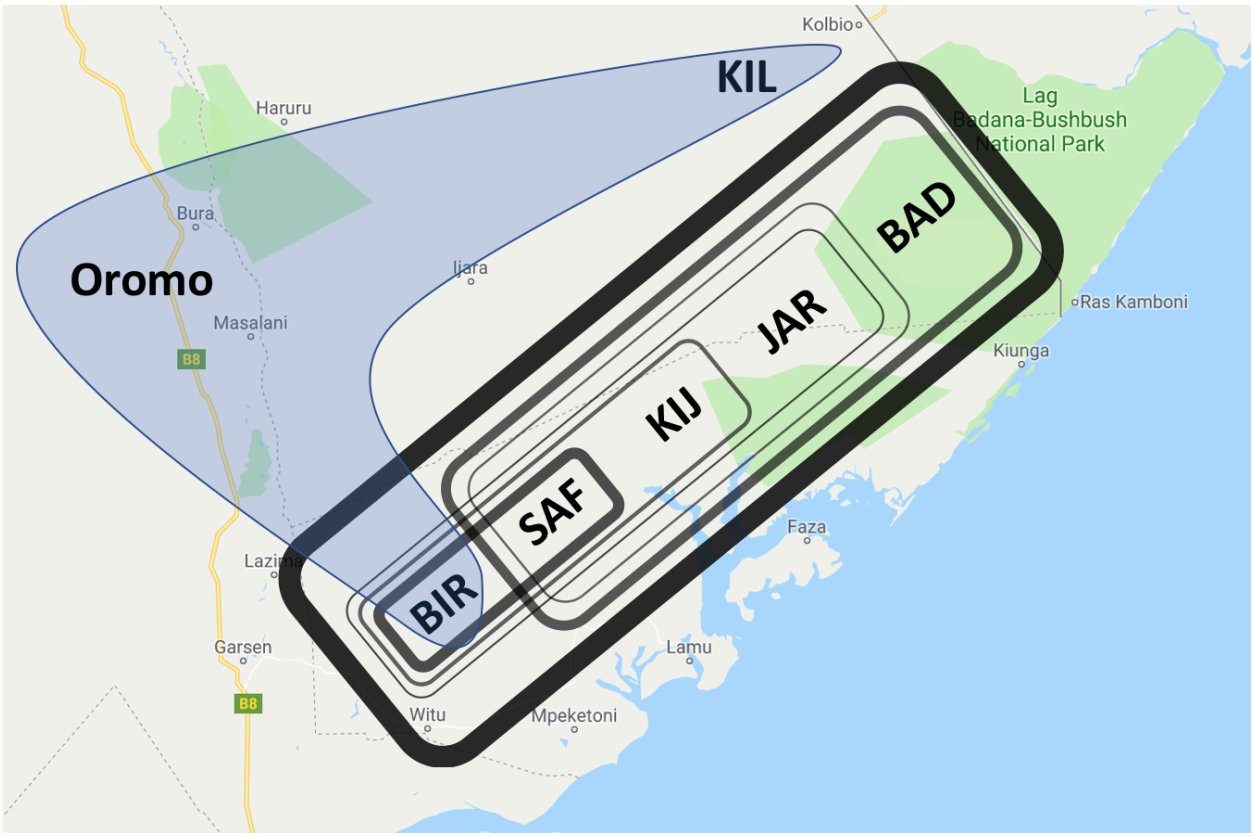
Figure 5: Glottometric diagram of the Boni dialects



The dialects represented are Bireeri (BIR), Safaree (SAF), Kijee (KIJ), Jara (JAR), Baddey (BAD) and Kilii (KIL). On the top left is Oromo, represented here because it exerts an influence on Bireeri and Kilii especially. The blue patch extending from Oromo represents the influence which has extended to affect the peripheral Boni dialects Kilii and Bireeri especially strongly, and is an *ad hoc* addition to the diagram.

If the dialects are displayed on a map true to their geographic locations, the resulting visualization is called a glottometric map instead. The configuration of Boni dialects in a chain along the coast makes it relatively easy to transform the glottometric diagram into a glottometric map, shown in Figure 9:

Figure 6: Glottometric map of the Boni dialects (map data copyright of Google 2018)



3.3 Discussion

A few observations can be made about this glottometric diagram (and map). First of all, the thickest and darkest bundle of isoglosses separates Kilii from the remaining dialects. This is the only subgroup for which the available linguistic (and historical) evidence of a split is convincing, and this is reflected by the thickness and darkness of the line separating Kilii from the rest. Second, the lines which surround various subgroups of remaining dialects are rather light and thin. They all appear to be of a similar darkness because they all have cohesiveness scores in a narrow range between 0.59 - 0.7. This reflects the fact that the evidence for any one subgroup is quite slim, and any proposed subgroup can be undermined with other data. They all receive approximately equal levels of support from the evidence provided. Instead of making an arbitrary choice as to which set of evidence defines a "true subgroup" and which are "contact effects", this can be understood as the normal situation for a group of closely related dialects which remain in contact.

Much information that would be lost in a tree diagram is retained in this glottometric diagram. For instance, the close relationship of Bireeri and Safaree is obscured by the fact that Heine judges Bireeri to have branched off quite early. In the glottometric diagram, the dual nature of Safaree is clear. The outline of what Heine calls the Central Convergence Area (Safaree, Kijee, Jara, Baddey) is clearly visible in the diagram. Another quite beautiful pattern which can be observed is the influence spreading out from Bireeri. Heine himself mentions that there seems to be a set of innovations spreading from Bireeri which affect each

successive neighbor less strongly. This pattern can clearly be seen in the glottometric diagram, where Bireeri has a series of lines surrounding it which grow thinner and lighter with distance. This illustrates very nicely the metaphor so often used by the wave model of linguistic change as ripples spreading in a pond. The subgroups proposed by Heine of South Aweer and Central Aweer are not surrounded by a prominent line, because the evidence supporting them is quite weak. Finally, Oromo was placed apart from the Boni dialects in a different shade, in order to represent the common innovations in Bireeri and Kilii which are due to both being in contact with Oromo (rather than directly in contact with each other). There is no convention for this including this type of information in a glottometric diagram yet.

One potentially misleading aspect of the glottometric diagram is that Kilii does not seem to be joined to the other dialects by any bundle of isoglosses at all; this is because the object of this paper is not to establish the innovations which characterize all the Boni dialects, but rather those which characterize subgroups within Boni. With a bit of searching, it should be easy to find many innovations exclusively shared by Boni as a unit in contrast to the other Sam languages Somali and Rendille, and hence to establish support for such a bundle of isoglosses.

4. **Conclusion**

In this paper, I first reviewed the evidence which Heine presents in support of his classification of Boni dialects. Two of the three proposed subgroups did not reach the level of support required to establish a secure genetic subgroup, so I put forth an alternative scenario to Heine's. I propose that Boni only underwent one split (between Kilii and the rest of the dialects) followed by *in situ* differentiation of the other dialects. The latter situation is an example of non-tree-like language differentiation, which has left a distinct pattern of overlapping innovation in the Boni dialects and is not suitably represented by a tree diagram. In order to represent it, I chose a glottometric diagram, which has certain advantages over the tree diagram.

First, building the glottometric diagram does not require a strict division into vertical and horizontal transmission, which is always fraught with difficulty in a closely knit dialect group like Boni. Second, it retains much of the information which is thrown out by a tree diagram. It permits certain observations to be made which would be otherwise impossible with a tree diagram. Third, it provides a much more natural and realistic account of the current observed pattern of innovations. The tree model forces one to make unnecessarily strong predictions about splits and contact, when the evidence may not justify such inferences.

There are a few areas in which further progress can be made. To an audience familiar with tree diagrams, a new method of visualization can seem difficult to interpret. However, the problems with the tree diagram, especially in dialect chains and linkages, are too numerous to ignore, and there is a need for a new method of visualization. A practical matter which needs to be addressed is how to make the information about subgroupiness and cohesiveness as salient as possible to the reader; the thickness and darkness conventions used in this paper could continue to be improved. Additionally, is still quite problematic to construct a glottometric diagram where too much information is missing from the database. It is hoped that future iterations of HG will address some of these remaining issues.

Abbreviations

BAD	Baddey
BIR	Bireeri
CA	Central Aweer
CCA	Central Convergence Area
HG	Historical Glottometry
JAR	Jara
KIJ	Kijee
KIL	Kilii
LDAK	Language and Dialect Atlas of Kenya
PB	Proto-Boni
SA	South Aweer
SAF	Safaree
SCA	Southern Convergence Area
SCCA	South-Central Convergence Area

References

Anttila, Raimo. 1989. *Historical and Comparative Linguistics.* Amsterdam: John Benjamins.

Bryant, David & Vincent Moulton. 2003. "Neighbor-Net: An Agglomerative Method for the Construction of Phylogenetic Networks" in *Molecular Biology and Evolution* 21 (2): 255-265.

Bryant, David, Flavia Filimon & Russel D. Gray. 2005. "Untangling our past: languages, trees, splits and networks." in *The evolution of cultural diversity: phylogenetic approaches* 67-83. London: UCL Press.

François, Alexandre. 2017. "Méthode comparative et chaînages linguistiques: Pour un modèle diffusionniste en généalogie des langues." in Jean-Léo Léonard (ed.): *Diffusion: implantation, affinités, convergence.* Mémoires de la Société de Linguistique de Paris, XXIV: 43-82. Louvain: Peeters.

François, Alexandre. 2014. "Trees, waves and linkages: models of language diversification." in Claire Bowern & Bethwyn Evans (eds.): *The Routledge Handbook of Historical Linguistics*: 161-189. New York: Routledge.

Heine, Bernd. 1977. "Bemerkungen zur Boni-Sprache (Kenia)" in *Afrika und Übersee LX*. 242-295.

Heine, Bernd. 1978. "The Sam Languages: A History of Rendille, Boni and Somali" in *Afroasiatic Linguistics* 6. 1-93.

Heine, Bernd & W.J.G. Möhlig. 1982. "Boni Dialects" in *Language and Dialect Atlas of Kenya:* Vol. X. Berlin: Dietrich Reimer.

Kalyan, Siva & Alexandre François. 2018. "Freeing the Comparative Method from the tree model: A framework for Historical Glottometry" in R. Kikusawa & L. Reid (eds.): *Let's talk about trees: Tackling problems in representing phylogenic relationships among languages*. Osaka: National Museum of Ethnology. 59-89.

Meinhof, Carl. 1912. *Die Sprachen der Hamiten*. Hamburg: Abhanglungen des Hamburgischen Kolonialinstituts.

Ross, Malcolm. 1988. *Proto-Oceanic and the Austronesian languages of western Melanesia*. Canberra: Pacific Linguistics.

Sasse, Hans-Jürgen. 1979. "Entlehnung morphophonemischer Regeln im Boni" in *Sprache und Geschichte in Afrika* 1. 93-108.

Schmidt, Johannes. 1872. *Die Verwandtschaftsverhältnisse der indogermanischen Sprachen.* Weimar: Hermann Böhlau.

Simons, Gary F. & Charles D. Fennig (eds.). 2017. *Ethnologue: Languages of the World, Twentieth Edition.* Dallas, Texas: SIL International. Online version: http://www.ethnologue.com.

Stiles, Daniel. 1988. "Historical interrelationships of the Boni with pastoral peoples of Somalia and Kenya", in *Kenya Past and Present* 20. 38-45.